Supplementary Materials of the Paper titled

"Estimation of Piecewise Continuous Regression Function in Finite Dimension using Oblique-axis Regression Tree with Applications in Image Denoising"

Subhasish Basak[1], Anik Roy[2], Partha Sarathi Mukherjee[1*]

[1]*Indian Statistical Institute, Kolkata, India*
[2]*Emory University, Atlanta, USA*

[*]Corresponding Author

S.1. Supplementary Materials

We provide the proofs of the theoretical results in this document.

S.1.1 Proof of Property 1

From **Eqn. 2.3**, we have

$$
\begin{aligned}
&\widehat{\Delta}(\mathcal{N}, \boldsymbol{\alpha}, c)\\
&= \frac{1}{|\mathcal{N}|}\left(\sum_{\boldsymbol{z}\in\mathcal{N}} w(\boldsymbol{z})^2 - \sum_{\boldsymbol{z}\in\mathcal{N}_L} w(\boldsymbol{z})^2 - \sum_{\boldsymbol{z}\in\mathcal{N}_R} w(\boldsymbol{z})^2 - |\mathcal{N}|\bar{w}^2 + |\mathcal{N}_R|\bar{w}_R^2 + |\mathcal{N}_L|\bar{w}_L^2\right)\\
&= \frac{1}{|\mathcal{N}|}\left(|\mathcal{N}_R|\bar{w}_R^2 + |\mathcal{N}_L|\bar{w}_L^2 - |\mathcal{N}|(\frac{|\mathcal{N}_R|\bar{w}_R}{|\mathcal{N}|} + \frac{|\mathcal{N}_L|\bar{w}_L}{|\mathcal{N}|})^2\right)\\
&= \frac{1}{|\mathcal{N}|^2}\left((|\mathcal{N}||\mathcal{N}_R| - |\mathcal{N}_R|^2)\bar{w}_R^2 + (|\mathcal{N}||\mathcal{N}_L| - |\mathcal{N}_L|^2)\bar{w}_L^2 - 2|\mathcal{N}_R||\mathcal{N}_L|\bar{w}_L\bar{w}_R\right)
\end{aligned}
$$

$$= \frac{|\mathcal{N}_L||\mathcal{N}_R|}{|\mathcal{N}|^2}\Big(\bar{w}_L - \bar{w}_R\Big)^2.$$

S.1.2 Proof of Theorem 2

Since $\mathcal{L}_n$ are leaf nodes, we have $\widehat{\Delta}(\mathcal{L}_n) \leq r_n$.

Now, consider any $(\boldsymbol{\alpha}, c)$ that splits $\mathcal{L}_n$ in half. Then, using **Eqn. 3.5** for that split, we have: $r_n \geq \frac{|\mathcal{L}_n^L||\mathcal{L}_n^R|}{|\mathcal{L}_n|^2}\Big(\bar{w}_L - \bar{w}_R\Big)^2$.

We know $\frac{|\mathcal{L}_n^L||\mathcal{L}_n^R|}{|\mathcal{L}_n|}\Big(\bar{w}_L - \bar{w}_R\Big)^2$ follows noncentral $\chi_1^2(k)$ with a non-centrality parameter which we call k. Therefore, we have

$$r_n \geq \frac{\chi_1^2(k)}{|\mathcal{L}_n|} \implies |\mathcal{L}_n| \geq \frac{\chi_1^2(k)}{r_n}.$$

As $\chi_1^2(k)$ is stocahastically larger than χ_1^2, we can say:

$$\mathbb{P}(|\mathcal{L}_n| \leq M) \leq \mathbb{P}(\chi_1^2 \leq M r_n).$$

As $r_n \to 0$, we can conclude that $|\mathcal{L}_n| \to \infty$ in probability. If we choose r_n such that the RHS of the above inequality becomes summable, then we will have almost sure convergence as well. In that case, we use a bound given by Ghosh (2021):

$$\begin{aligned}\mathbb{P}(\chi_1^2 \leq Mr_n) &\leq exp\left[\frac{1}{2}\left(1 - Mr_n + log(Mr_n)\right)\right] \\ &= \sqrt{Mr_n} exp\left(\frac{1 - Mr_n}{2}\right) \leq \sqrt{eMr_n}.\end{aligned}$$

Therefore, if $\{\sqrt{r_n}\}$ is summable, we have $|\mathcal{L}_n| \to \infty$ almost surely.

S.1.3 Proof of Theorem 3

Suppose, $\mu(\mathcal{L}_n) \to \ell$. Since $|\mathcal{L}_n| \approx n^p \mu(\mathcal{L}_n)$, for large n, we also have $|\mathcal{L}_n| \to \infty$. Assume $\bigcap_n \mathcal{L}_n = \mathcal{L}$. As $\mu(\mathcal{L}) = \ell > 0$, we have $\mathcal{L}_n \to \mathcal{L}$ which is non empty. Now consider a split along the direction $\boldsymbol{\alpha}$ and constant c, which splits $\mathcal{L}$ into $\mathcal{L}^L$ and $\mathcal{L}^R$. The same hyperplane also splits $\mathcal{L}_n$ in $\mathcal{L}_n^L$ and $\mathcal{L}_n^R$. Then, we must have $\mathcal{L}_n^L \to \mathcal{L}^L$ and $\mathcal{L}_n^R \to \mathcal{L}^R$. As $\mathcal{L}_n$ are leaf nodes, we must have $\widehat{\Delta}(\mathcal{L}_n, \boldsymbol{\alpha}, c) \leq r_n$. Therefore,

$$\begin{aligned}&\frac{1}{|\mathcal{L}_n|}\left[\sum_{\boldsymbol{z}\in\mathcal{L}_n}\big(w(x,y) - \bar{w}\big)^2 - \sum_{\boldsymbol{z}\in\mathcal{L}_n^R}\big(w(x,y) - \bar{w}_R\big)^2 - \sum_{\boldsymbol{z}\in\mathcal{L}_n^L}\big(w(x,y) - \bar{w}_L\big)^2\right] \leq r_n \\ &\implies \frac{|\mathcal{L}_n^L||\mathcal{L}_n^R|}{|\mathcal{L}_n|^2}\left(\bar{w}_L - \bar{w}_R\right)^2 \leq r_n. \\ &\implies \frac{\mu(\mathcal{L}_n^L)\mu(\mathcal{L}_n^R)}{\mu(\mathcal{L}_n)^2}\left(\bar{w}_L - \bar{w}_R\right)^2 \leq r_n.\end{aligned}$$

Note that $\bar{w}_R$ and $\bar{w}_L$ are the averages of right and left child respectively.

Now observe:

$$
\begin{aligned}
\bar{w}_L &= \sum_{\boldsymbol{z}\in\mathcal{L}_n^L} w(\boldsymbol{z})/|\mathcal{L}_n^L| \\
&= \sum_{\boldsymbol{z}\in\mathcal{L}_n^L} (f(\boldsymbol{z})+\epsilon(\boldsymbol{z}))/|\mathcal{L}_n^L| \\
&= \frac{\frac{1}{n^p}\sum_{\boldsymbol{z}\in\mathcal{L}_n^L} f(\boldsymbol{z})}{\frac{|\mathcal{L}_n^L|}{n^p}} + \frac{\sum_{\boldsymbol{z}\in\mathcal{L}_n^L}\epsilon(\boldsymbol{z})}{|\mathcal{L}_n^L|} \\
&\to \left(\int_{\mathcal{L}^L} f(\boldsymbol{z})d\mu\right)/\mu(\mathcal{L}^L).
\end{aligned}
$$

As the first term is the Riemann sum of the integral of $f(\boldsymbol{z})\mathbb{I}_{\mathcal{L}_n^L}(\boldsymbol{z})$, which is bounded by f, we can use Dominated Convergence Theorem (DCT) to get that limit. The second term converges to 0, due to the Strong Law of Large Numbers (SLLN). As $r_n \to 0$, combining the above two results, we have:

$$
\frac{\mu(\mathcal{L}^L)\mu(\mathcal{L}^R)}{\mu(\mathcal{L})^2}\left(\frac{\int_{\mathcal{L}^L} f(\boldsymbol{z})d\mu}{\mu(\mathcal{L}^L)} - \frac{\int_{\mathcal{L}^R} f(\boldsymbol{z})d\mu}{\mu(\mathcal{L}^R)}\right)^2 = 0. \tag{S.1.1}
$$

As $\mu(\mathcal{L}) \neq 0$, and the choices of α and c were arbitrary, we note that:

$$
\begin{aligned}
\frac{\int_{\mathcal{L}^L} f(\boldsymbol{z})d\mu}{\mu(\mathcal{L}^L)} &= \frac{\int_{\mathcal{L}^R} f(\boldsymbol{z})d\mu}{\mu(\mathcal{L}^R)} \\
\implies \frac{\int_{\mathcal{L}^L} f(\boldsymbol{z})d\mu}{\mu(\mathcal{L}^L)} &= \frac{\int_{\mathcal{L}} f(\boldsymbol{z})d\mu}{\mu(\mathcal{L})} \\
\implies \frac{\int_{\boldsymbol{\alpha}^T z\le c} f(\boldsymbol{z})d\mu}{\mu(\{\boldsymbol{\alpha}^T z \le c\}\bigcap\mathcal{L})} &= \frac{\int_{\mathcal{L}} f(\boldsymbol{z})d\mu}{\mu(\mathcal{L})}
\end{aligned} \tag{S.1.2}
$$

for all possible splits. Define $h(\boldsymbol{z}) = \frac{f(\boldsymbol{z})}{\int_{\mathcal{L}} f(\boldsymbol{z})d\mu}$. Consider a random variable

$\boldsymbol{X}$ which takes values inside $\mathcal{L}$ with density $h(z)$. From Eqn. (S.1.2), we see that:

$$\int_{\boldsymbol{\alpha}^T z \leq c} h(\boldsymbol{z}) d\mu = \frac{\mu(\{\boldsymbol{\alpha}^T z \leq c\} \bigcap \mathcal{L})}{\mu(\mathcal{L})} = \mathbb{P}(\boldsymbol{\alpha}^T \boldsymbol{X} \leq c).$$

Therefore, all linear combinations of $\boldsymbol{X}$ have the same distribution as linear combinations of a uniform random variable on $\mathcal{L}$. Hence, by Cramér-Wold theorem, we can say $\boldsymbol{X}$ has uniform distribution on $\mathcal{L}$, i.e., $h(z) = \frac{1}{\mu(\mathcal{L})}$ $a.e. \implies f(\boldsymbol{z}) = \frac{\int_{\mathcal{L}} f(\boldsymbol{z}) d\mu}{\mu(\mathcal{L})}$ $a.e.$ Hence $f(\boldsymbol{z})$ is $a.e.$ constant on $\mathcal{L}$.

S.1.4 Proof of Corollary 1

As $\mathcal{L}_n$ are convex, so is $\mathcal{L}$. Therefore, if $\mu(\mathcal{L}) = 0$, then it is a convex set of dimension less than p.

Suppose, the dimension of $\mathcal{L}$ is $k < p$. Consider μ_k to be the Lebesgue measure corresponding to $\mathbb{R}^k$ on $\mathcal{L}$. Hence, we have $\mu_k(\mathcal{L}) \neq 0$.

Now as $\mathcal{L}$ is of dimension k, there must exist $(n-k)$ axes which are not parallel to the k-dimensional affine subspace $\bar{L}$ which contains $\mathcal{L}$. Call them $e_1, e_2, \ldots e_{n-k}$. Let $\mathcal{R}_{n-k}$ be the subspace generated by $e_1, e_2, \ldots e_{n-k}$. Now consider $\mathcal{R}_x = \boldsymbol{z_x} + \mathcal{R}_{n-k}$, where z_x is the closest lattice point to $\boldsymbol{x}$, and define a new function:

$$J_n(\boldsymbol{x}) = \frac{\sum\limits_{y \in \mathcal{R}_x \bigcap \mathcal{L}_n} w(\boldsymbol{y})}{|\mathcal{R}_x \bigcap \mathcal{L}_n|} \quad \forall \boldsymbol{x} \in \bar{L}.$$

We are basically superimposing the whole image onto $\bar{L}$. Now, if we define $\Delta_J(\mathcal{L}_n)$ in a similar fashion to Δ, it is easy to observe that $\Delta_J(\mathcal{L}_n) \leq \Delta(\mathcal{L}_n) \leq r_n$. Proceeding similarly to the proof of the above theorem, we get:

$$\frac{\mu_k(\bar{\mathcal{L}}_n^L)\mu_k(\bar{\mathcal{L}}_n^R)}{\mu(\bar{\mathcal{L}}_n)^2}\left(\bar{J}_n^L - \bar{L}_n^R\right)^2 \leq r_n,$$

where

$$\begin{aligned} \bar{J}_n^L &= \sum_{\boldsymbol{x} \in \bar{\mathcal{L}}_n^L} J_n(\boldsymbol{x})/|\bar{\mathcal{L}}_n^L| \\ &= \int\limits_{\bar{\mathcal{L}}_n^L} J_n(\boldsymbol{x}) d\mu_k x/(|\bar{\mathcal{L}}_n^L|/n^k), \end{aligned}$$

and $\bar{J}_n^R$ is defined in a similar fashion. Observe that if f is continuous on $\mathcal{L}$, $J_n(\boldsymbol{x})$ converges point-wise to $f(x)$. Using dominated convergence theorem (DCT) on the integral, we get:

$$\int\limits_{\bar{\mathcal{L}}_n^L} J_n(\boldsymbol{x}) d\mu_k x \to \int\limits_{\mathcal{L}^L} f(\boldsymbol{x}) d\mu_k x.$$

Then, proceeding similarly as the proof of **Theorem 3**, we get f is a.e. constant in $\mathcal{L}$. Remember that for small enough $\mu(\mathcal{L}_n)$, we assume only one JLC is present in $\mathcal{L}_n$. Hence, if f is not continuous on a measure set with respect to μ_k on $\mathcal{L}$, the same result holds. Otherwise, because of the assumption of a JLC being approximately a lower-dimensional plane, and $\mathcal{L}$ being convex, it must mean that the JLC contains $\mathcal{L}$ entirely or $\mathcal{L}$ contains the JLC entirely.

S.1.5 Proof of Corollary 2

Applying $k = 1$ in **Corollary 1**.

S.1.6 Proof of Property 4

Property 4 holds because at most countably many sets of these unions are distinctly non-empty, as there are at most countably many JLCs.

S.1.7 Proof of Lemma 1

Suppose that for all n, $\mathcal{L}_n^{\boldsymbol{x}}$ contains at least one discontinuity point. Consider $\mathcal{N}_\epsilon(\boldsymbol{x})$ to be an ϵ ball around $\boldsymbol{x}$. As $\gamma(\mathcal{L}_n^{\boldsymbol{x}}) \to 0$, we can say that for every ϵ, there exists an $n_\epsilon > 0$ such that $\forall\, m > n_\epsilon$, $\gamma(\mathcal{L}_n^{\boldsymbol{x}}) < \epsilon$, which implies $\mathcal{L}_m^{\boldsymbol{x}} \subseteq \mathcal{N}_\epsilon(\boldsymbol{x})$. However, as $\mathcal{L}_m^{\boldsymbol{x}}$ contains at least one discontinuity point, say $\boldsymbol{y_n}$. Since $\{\boldsymbol{y_n}\}$ is a bounded sequence, there exists a convergent subsequence. Assume that the convergent subsequence converges to $\boldsymbol{y}$. As $\boldsymbol{y} \in \overline{\bigcap_n \mathcal{L}_n^{\boldsymbol{x}}}$, we have $\|\boldsymbol{y} - \boldsymbol{x}\| = 0 \implies \boldsymbol{y} = \boldsymbol{x}$. Since the set of jump points is closed by definition, either $\boldsymbol{y}$ is a discontinuity point or a singular point. Both options contradict our assumption that $\boldsymbol{x}$ is a non-singular continuity point. Therefore, $\mathcal{L}_n^{\boldsymbol{x}}$ cannot contain a discontinuity point $\forall\ n$. Since $\mathcal{L}_n^{\boldsymbol{x}}$ is a non-increasing sequence of sets, we conclude that after some N, it does not contain any discontinuity point.

S.1.8 Proof of Theorem 5

We prove this only on continuity points for which $\mathcal{L}^{\boldsymbol{x}}$ *does not contain an entire JLC* as the remaining points have measure zero. We divide this proof into three different cases:

Case 1: f is not *a.e.* constant in $\bigcap_n \mathcal{L}_n^{\boldsymbol{x}}$. Hence we also have $\lim_{n\to\infty} \mu(\mathcal{L}_n^{\boldsymbol{x}}) = 0$. As per the assumptions, $\bigcap_n \mathcal{L}_n^{\boldsymbol{x}}$ cannot lie entirely on a JLC. Hence, we have $\lim_{n\to\infty} \gamma(\mathcal{L}_n^{\boldsymbol{x}})$. Using **Lemma 1**, $\exists\ N$ such that $\mathcal{L}_n^{\boldsymbol{x}}$ does not contain any JLC $\forall\ n > N$. In this case, $\forall \boldsymbol{z} \in \mathcal{L}_n^{\boldsymbol{x}}$, we have:

$$
\begin{aligned}
& |w(\boldsymbol{z}) - f(\boldsymbol{x})| \leq |f(\boldsymbol{z}) - f(\boldsymbol{x})| + |\varepsilon(\boldsymbol{z})| \leq C_\ell \gamma((\mathcal{L}_n^{\boldsymbol{x}})) + |\varepsilon(\boldsymbol{z})| \\
\implies & \left| \sum_{z \in \mathcal{L}_n^{\boldsymbol{x}}} w(\boldsymbol{z}) - f(x) \right| \leq |\mathcal{L}_n^{\boldsymbol{x}}| C_\ell \gamma(\mathcal{L}_n^{\boldsymbol{x}}) + \left| \sum_{z \in \mathcal{L}_n^{\boldsymbol{x}}} \varepsilon(\boldsymbol{z}) \right| \\
\implies & \left| \widehat{f}_n(\boldsymbol{x}) - f(x) \right| \leq C_\ell \gamma(\mathcal{L}_n^{\boldsymbol{x}}) + \left| \frac{\sum_{z \in \mathcal{L}_n^{\boldsymbol{x}}} \varepsilon(\boldsymbol{z})}{|\mathcal{L}_N^{\boldsymbol{x}}|} \right| \qquad \text{(S.1.3)} \\
\implies & \left| \widehat{f}_n(\boldsymbol{x}) - f(x) \right| \xrightarrow{n\to\infty} 0, \text{ using SLLN and } \textbf{Corollary 1}.
\end{aligned}
$$

Case 2: f is *a.e.* constant in $\bigcap_n \mathcal{L}_n^{\boldsymbol{x}}$ and $\lim_{n\to\infty} \mu(\mathcal{L}_n^{\boldsymbol{x}}) \neq 0$.

In this case, we know that f is *a.e.* constant in $\bigcap_n \mathcal{L}_n^{\boldsymbol{x}}$. So we have:

$$\begin{aligned}\widehat{f}_n(\boldsymbol{x}) - f(x) &= \frac{\sum_{z\in\mathcal{L}_n^{\boldsymbol{x}}} (f(\boldsymbol{z}) - f(\boldsymbol{x}))}{|\mathcal{L}_n^{\boldsymbol{x}}|} + \frac{\sum_{z\in\mathcal{L}_n^{\boldsymbol{x}}} \varepsilon(\boldsymbol{z})}{|\mathcal{L}_n^{\boldsymbol{x}}|} \\ &= \frac{\frac{1}{n^p}\sum_{z\in\mathcal{L}_n^{\boldsymbol{x}}} (f(\boldsymbol{z}) - f(\boldsymbol{x}))}{|\mathcal{L}_n^{\boldsymbol{x}}|/n^p} + \frac{\sum_{z\in\mathcal{L}_n^{\boldsymbol{x}}} \varepsilon(\boldsymbol{z})}{|\mathcal{L}_n^{\boldsymbol{x}}|}.\end{aligned}$$

The numerator of the first term converges to $\int_{\bigcap_n \mathcal{L}_n^{\boldsymbol{x}}} (f(\boldsymbol{z}) - f(\boldsymbol{x}))$ which is 0 as f is almost surely *a.e.* constant there, and the denominator converges to $\lim_{n\to\infty} \mu(\mathcal{L}_n^{\boldsymbol{x}})$ which is non-zero. Therefore, we have $\lim_{n\to\infty} \widehat{f}_n(\boldsymbol{x}) = f(\boldsymbol{x})$ a.s.

Case 3: f is *a.e.* constant on $\bigcap_n \mathcal{L}_n^{\boldsymbol{x}}$ and $\lim_{n\to\infty} \mu(\mathcal{L}_n^{\boldsymbol{x}}) = 0$.

Observe that if $\exists\ k$ such that $\lim_{n\to\infty} \mu_k(\mathcal{L}_n^{\boldsymbol{x}}) \neq 0$, we can proceed similar to Case 2. Therefore, without any loss of generality, we can assume $\gamma(\mathcal{L}_n^{\boldsymbol{x}}) \to 0$, and then proceed similarly to Case 1.

From above cases, we conclude that $\lim_{n\to\infty} \widehat{f}_n(\boldsymbol{x}) = f(\boldsymbol{x})$ *a.e.*, almost surely.

S.1.9 Proof of Theorem 6

While the following diagram of $\mathcal{L}_n$ is drawn for the 2-D case, the general p-dimensional situation can be understood similarly. Note that the actual leaf node is a convex polygon in the 2-D case and a convex polyhedron in the general p-dimensional scenario. For the simplicity of drawing, the shape is shown as an ellipse rather than a convex polygon with many sides.

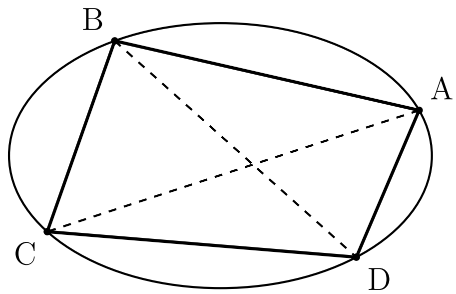

Diagram showing the BD-axis and AC-hyperplane of a leaf node $\mathcal{L}_n$.

The line BD represents the maximum possible distance in the set along a given direction $\boldsymbol{u}$. AC is perpendicular to the gradient $f^{'}(\boldsymbol{x}_0)$, and hence, for the general p-dimension, it is a hyperplane. Next, we want to split the line BD along AC. In the general situation, we can split it using the plane $(f^{'}(\boldsymbol{x}_0))^t(\boldsymbol{x} - \boldsymbol{x}_0) = 0$. The intuition behind this is as follows. As the gradient along AC is close to zero, we can ignore changes in $f(x)$ along the line BD. Moreover, we can move the line or hyperplane in such a manner that it splits the set into two equal parts, which lets us equate the fraction $\frac{n_1 n_2}{|\mathcal{L}_n|^2}$ to $\frac{1}{4}$.

Without any loss of generality, we can assume that $\boldsymbol{x_0}$ is the point of intersection. If it is not, we can work with the intersection point itself, but the value of the function f at that point will be similar to the given point, as $|f^{'}(\boldsymbol{x})|$ lies between 3δ and δ. We then denote the two children of the leaf node $\mathcal{L}$ by A and B. Approximating the average intensity by the corresponding Riemann integral and applying the SLLN to the random errors, we have

$$\begin{aligned}
\bar{w}_1 &\approx \frac{\int\limits_A f(\boldsymbol{x})d\boldsymbol{x}}{\mu(A)} + O\left(\sqrt{\frac{\log\log|\mathcal{L}_n|}{|\mathcal{L}_n|}}\right) \\
&= \frac{1}{\mu(A)}\int\limits_A \Big(f(\boldsymbol{x_0}) + (f^{'}(\boldsymbol{\zeta})^t(\boldsymbol{x}-\boldsymbol{x_0})\Big)d\boldsymbol{x} + O\left(\sqrt{\frac{\log\log|\mathcal{L}_n|}{|\mathcal{L}_n|}}\right) \\
&= \frac{1}{\mu(A)}\int\limits_A \Big(f(\boldsymbol{x_0}) + (f^{'}(\boldsymbol{x_0})+\boldsymbol{v})^t(\boldsymbol{x}-\boldsymbol{x_0})\Big)d\boldsymbol{x} + O\left(\sqrt{\frac{\log\log|\mathcal{L}_n|}{|\mathcal{L}_n|}}\right) \\
&= \frac{1}{\mu(A)}\int\limits_A \Big(f(\boldsymbol{x_0}) + (f^{'}(\boldsymbol{x_0}))^t(\boldsymbol{x}-\boldsymbol{x_0})\Big)d\boldsymbol{x} + O\Big(|\boldsymbol{v}|\gamma(\mathcal{L}_n)\Big) + O\left(\sqrt{\frac{\log\log|\mathcal{L}_n|}{|\mathcal{L}_n|}}\right) \\
&= f(\boldsymbol{x_0}) + \frac{1}{\mu(A)}\int\limits_A (f^{'}(\boldsymbol{x_0}))^t(\boldsymbol{x}-\boldsymbol{x_0})d\boldsymbol{x} + O\Big(|\boldsymbol{v}|\gamma(\mathcal{L}_n)\Big) + O\left(\sqrt{\frac{\log\log|\mathcal{L}_n|}{|\mathcal{L}_n|}}\right) \\
&= f(\boldsymbol{x_0}) + \frac{1}{\mu(A)}\int\limits_A (f^{'}(\boldsymbol{x_0}))^t(\boldsymbol{x}-\boldsymbol{x_0})d\boldsymbol{x} + O(|\boldsymbol{v}|) + O\left(\sqrt{\frac{\log\log|\mathcal{L}_n|}{|\mathcal{L}_n|}}\right).
\end{aligned}$$

We have replaced $f^{'}(\zeta)$ with $\big(f^{'}(\boldsymbol{x}_0)+\boldsymbol{v}\big)$ in the above expressions. Therefore, $O(|\boldsymbol{v}|) = O(\delta)$, and hence we can choose δ arbitrarily small. Also, note that $f^{'}(\boldsymbol{x_0})^t(\boldsymbol{x}-\boldsymbol{x_0}) = |f^{'}(\boldsymbol{x_0})|H_{\boldsymbol{x}}$ where $H_{\boldsymbol{x}}$ is the height of the point $\boldsymbol{x}$ from the line or hyperplane AC, and $\mu(A) = \mu(B) = \frac{1}{2}\mu(\mathcal{L}_n)$. Moreover. if the intersected part of the hyperplane has measure M with respect to μ_{p-1}, then $2M\gamma(\mathcal{L}_n) \geq \mu(\mathcal{L}_n)$ implying that $\frac{1}{\mu(A)} \geq \frac{1}{M\gamma(\mathcal{L}_n)}$. Therefore,

$$\bar{f}_1 - \bar{f}_2 = \frac{2|f'(\boldsymbol{x_0})|}{\mu(\mathcal{L}_n)} \left(\int_A H_{\boldsymbol{x}} d\boldsymbol{x} + \int_B H_{\boldsymbol{x}} d\boldsymbol{x} \right) \geq \frac{|f'(\boldsymbol{x_0})|}{M\gamma(\mathcal{L}_n)} \left(\int_A H_{\boldsymbol{x}} d\boldsymbol{x} + \int_B H_{\boldsymbol{x}} d\boldsymbol{x} \right). \tag{S.1.4}$$

The above integrals are larger than the integrals restricted to the quadrilateral or polyhedron, and the integrated height of the polyhedron is given by $\frac{H^2 M}{p(p+1)}$. Hence, we have

$$\frac{n_1 n_2}{|\mathcal{L}_n|^2} \Big(\bar{w}_1 - \bar{w}_2 \Big)^2 + O(|\boldsymbol{v}|) + O\left(\frac{\log\log|\mathcal{L}_n|}{\sqrt{|\mathcal{L}_n|}} \right) > \frac{|f'(\boldsymbol{x_0})|^2}{4\gamma(\mathcal{L}_n)^2 p^2 (p+1)^2} (H_A^2 + H_B^2)^2$$
$$\geq \frac{|f'(\boldsymbol{x_0})|^2}{\gamma(\mathcal{L}_n)^2 p^2 (p+1)^2} \frac{(H_A + H_B)^4}{16} = K_p^{-1} \gamma(\mathcal{L}_n)^{-2} |f'(\boldsymbol{x_0})|^{-2} \Big(f'(\boldsymbol{x_0}))^t \vec{\gamma}_{\boldsymbol{u}}(\mathcal{L}_n) \Big)^4.$$

In the above expressions, $\vec{\gamma}_{\boldsymbol{u}}(\mathcal{L}_n)$ represents BD as shown in the above diagram, that is, the vector corresponding to the largest distance along the direction $\boldsymbol{u}$. As $\Delta \leq r_n$, we get

$$\Big(f'(\boldsymbol{x_0}))^t \vec{\gamma}_{\boldsymbol{u}}(\mathcal{L}_n) \Big)^4 \leq K_p r_n' \gamma(\mathcal{L}_n)^2 |f'(\boldsymbol{x_0})|^2$$
$$\Big(f'(\boldsymbol{x_0}))^t \vec{\gamma}_{\boldsymbol{u}}(\mathcal{L}_n) \Big) = O\Big(\big(r_n' \gamma(\mathcal{L}_n)^2 \big)^{\frac{1}{4}} |f'(\boldsymbol{x_0})|^{1/2} \Big), \tag{S.1.5}$$

where $r_n' = r_n + O(|\boldsymbol{v}|) + O\Big(\frac{\log\log|\mathcal{L}_n|}{\sqrt{|\mathcal{L}_n|}} \Big)$. We can ignore $O(|\boldsymbol{v}|) < \delta$ as we can choose it arbitrarily small. As $\gamma(\mathcal{L}_n)^2 \to 0$, we have

$$(f^{'}(\boldsymbol{x_0}))^t \overrightarrow{\gamma_{\boldsymbol{u}}}(\mathcal{L}_n)) = o\Big((r_n^{'})^{\frac{1}{4}}\Big). \tag{S.1.6}$$

Since $|\mathcal{L}_n| > \chi_1^2/r_n$, and $\frac{\log\log|\mathcal{L}_n|}{\sqrt{|\mathcal{L}_n|}}$ is a decreasing function in $|\mathcal{L}_n|$,

$$o\left(\frac{\log\log|\mathcal{L}_n|}{\sqrt{|\mathcal{L}_n|}}\right) \le o\left(\sqrt{r_n\Big|\log|\log r_n|\Big|}C_{\chi_1^2}\right).$$

Hence,

$$(f^{'}(\boldsymbol{x_0}))^t \overrightarrow{\gamma_{\boldsymbol{u}}}(\mathcal{L}_n) = o\left(\left(\sqrt{r_n\Big|\log|\log r_n|\Big|}C_{\chi_1^2}\right)^{1/4}\right). \tag{S.1.7}$$

Note that the order in the RHS of the Eqn. (S.1.7) is free of $\boldsymbol{u}$. Using this, we have

$$\begin{aligned}
\widehat{f}_n(\boldsymbol{x}_0) - f(\boldsymbol{x}_0) &= \frac{\sum\limits_{z\in\mathcal{L}_n^{\boldsymbol{x}_0}}(f(\boldsymbol{z}) - f(\boldsymbol{x_0}))}{|\mathcal{L}_n^{\boldsymbol{x}_0}|} + \frac{\sum\limits_{z\in\mathcal{L}_n^{\boldsymbol{x}_0}}\varepsilon(\boldsymbol{z})}{|\mathcal{L}_n^{\boldsymbol{x}_0}|} \\
&= o\left(\left(\sqrt{r_n\Big|\log|\log r_n|\Big|}C_{\chi_1^2}\right)^{1/4}\right) + o\left(\frac{\log\log|\mathcal{L}_n^{\boldsymbol{x}_0}|}{\sqrt{|\mathcal{L}_n^{\boldsymbol{x}_0}|}}\right) \\
&= o\left(\left(\sqrt{r_n\Big|\log|\log r_n|\Big|}C_{\chi_1^2}\right)^{1/4}\right) + o\left(\left(\sqrt{r_n\Big|\log|\log r_n|\Big|}C_{\chi_1^2}\right)\right) \\
&= o\left(\left(\sqrt{r_n\Big|\log|\log r_n|\Big|}C_{\chi_1^2}\right)^{1/4}\right).
\end{aligned} \tag{S.1.8}$$

The above expression provides the convergence rate of $\widehat{f}$ at the points where the first derivative is continuous and is non-zero. If the function f behaves like a constant locally, then the diameter of the leaf-node at those points can have a non-zero limit as $n \to \infty$. This is due to **Theorem 3**.

References

Ghosh, M. (2021). Exponential tail bounds for chisquared random variables. *Journal of Statistical Theory and Practice 15*(2), 35.